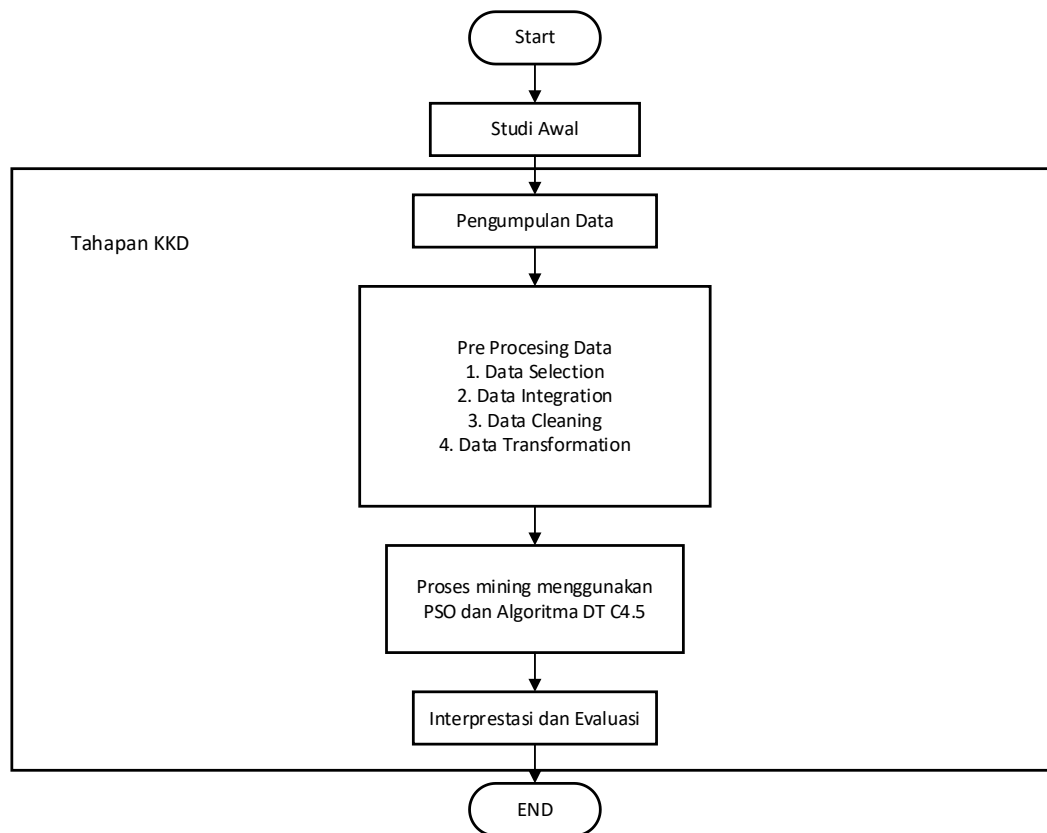


## BAB III

### METODE PENELITIAN

#### 3.1 Metode Penelitian

Pada bab ini akan membahas langkah-langkah dari proses penelitian yang akan dilaksanakan, dalam melakukan analisa dan mencari pola data untuk dijadikan sebuah dataset dalam memudahkan penelitian dan dapat berjalan dengan sistematis dan memenuhi tujuan yang diinginkan maka dibuat alur dalam tahapan penelitian yang akan dilakukan berikut :



Gambar 3. 1 Alur dalam Tahapan Penelitian

Tahapan pada gambar 3.1 . Adalah proses data mining pada penelitian ini :

1. Tahapan studi awal merupakan tahap awal dalam melakukan penelitian. Pada tahap ini, peneliti melakukan observasi atau pemahaman penelitian yang meliputi tujuan dan persyaratan proyek dengan jelas dalam hal bisnis atau unit penelitian secara keseluruhan, menterjemahkan tujuan dan batasan ke dalam perumusan definisi masalah data mining, menyiapkan strategi awal untuk mencapai tujuan tersebut.
2. Tahap pengumpulan data merupakan suatu tahapan pengumpulan data menggunakan analisis data eksplorasi untuk membiasakan diri dengan data dan menemukan wawasan awal dan mengevaluasi kualitas data. Data penelitian yang digunakan diperoleh dari situs kaggle.com (<https://www.kaggle.com/datasets/tejashvi14/employee-future-prediction>).
3. *Preprocessing* data merupakan proses pengolahan data mentah kedalam bentuk yang lebih mudah dipahami, proses ini penting dilakukan karena data mentah sering kali tidak memiliki format yang teratur.

a) Data Selection

Pada proses seleksi ini, dipilih dari keseluruhan data dari dataset Kaggle.com yaitu data *employee future prediction*. Dari hasil seleksi akan digunakan sebanyak 4.653 data karyawan untuk perhitungan data mining.

b) Data Integration

Proses ini dilakukan karena perlu adanya perubahan skema struktur data dari data mentah ke data yang siap akan dilakukan proses *preprocessing* lebih lanjut dengan tahap *cleaning*.

	A	B	C	D	E	F	G	H	I	J	K
1	Education	JoiningYear	City	PaymentTier	Age	Gender	EverBenched	ExperienceInCurrentDomain	LeaveOrNot		
2	Bachelors	2017	Bangalore	3	34	Male	No	0	0		
3	Bachelors	2013	Pune	1	28	Female	No	3	1		
4	Bachelors	2014	New Delhi	3	38	Female	No	2	0		
5	Masters	2016	Bangalore	3	27	Male	No	5	1		
6	Masters	2017	Pune	3	24	Male	Yes	2	1		
7	Bachelors	2016	Bangalore	3	22	Male	No	0	0		
8	Bachelors	2015	New Delhi	3	38	Male	No	0	0		
9	Bachelors	2016	Bangalore	3	34	Female	No	2	1		
10	Bachelors	2016	Pune	3	23	Male	No	1	0		
11	Masters	2017	New Delhi	2	37	Male	No	2	0		
12	Masters	2012	Bangalore	3	27	Male	No	5	1		
13	Bachelors	2016	Pune	3	34	Male	No	3	0		
14	Bachelors	2018	Pune	3	32	Male	Yes	5	1		
15	Bachelors	2016	Bangalore	3	39	Male	No	2	0		
16	Bachelors	2012	Bangalore	3	37	Male	No	4	0		
17	Bachelors	2017	Bangalore	1	29	Male	No	3	0		
18	Bachelors	2014	Bangalore	3	34	Female	No	2	0		
19	Bachelors	2014	Pune	3	34	Male	No	4	0		
20	Bachelors	2015	Pune	2	30	Female	No	0	1		
21	Bachelors	2016	New Delhi	2	22	Female	No	0	1		

Gambar 3. 2 *Data Sebelum Integration*

Data akan berubah skema atau struktur menjadi seperti pada Gambar 3.3 yang mudah untuk diproses. Data tabel didefinisikan memanjang ke samping kanan dengan penggunaan kolom yang lebih banyak dan sesuai secara rapi serta type file yang sebelumnya .csv di export menjadi type data excel sehingga memudahkan untuk merubah data.

	A	B	C	D	E	F	G	H	I
1	Education	JoiningYear	City	PaymentTier	Age	Gender	EverBenched	ExperienceInCurrentDomain	LeaveOrNot
2	Bachelors	2017	Bangalore	3	34	Male	No	0	0
3	Bachelors	2013	Pune	1	28	Female	No	3	1
4	Bachelors	2014	New Delhi	3	38	Female	No	2	0
5	Masters	2016	Bangalore	3	27	Male	No	5	1
6	Masters	2017	Pune	3	24	Male	Yes	2	1
7	Bachelors	2016	Bangalore	3	22	Male	No	0	0
8	Bachelors	2015	New Delhi	3	38	Male	No	0	0
9	Bachelors	2016	Bangalore	3	34	Female	No	2	1
10	Bachelors	2016	Pune	3	23	Male	No	1	0
11	Masters	2017	New Delhi	2	37	Male	No	2	0
12	Masters	2012	Bangalore	3	27	Male	No	5	1
13	Bachelors	2016	Pune	3	34	Male	No	3	0
14	Bachelors	2018	Pune	3	32	Male	Yes	5	1
15	Bachelors	2016	Bangalore	3	39	Male	No	2	0
16	Bachelors	2012	Bangalore	3	37	Male	No	4	0
17	Bachelors	2017	Bangalore	1	29	Male	No	3	0
18	Bachelors	2014	Bangalore	3	34	Female	No	2	0
19	Bachelors	2014	Pune	3	34	Male	No	4	0
20	Bachelors	2015	Pune	2	30	Female	No	0	1
21	Bachelors	2016	New Delhi	2	22	Female	No	0	1

Gambar 3. 3 Data Sesudah *Integration*

c) Data Cleaning

Pada proses ini data yang tidak perlu dipakai pada penggunaan pengolahan data dan data yang akan menyebabkan kekacauan data sehingga data menyimpang atau tidak konsisten.

Pada dataset ini tidak perlu dilakukan proses *cleaning* dikarenakan tidak ada data yang *missing* sehingga data siap di proses.

d) Data *Transformation*

Proses *Transformation* merupakan perubahan data tahap akhir yang siap diolah dan dimasukkan kedalam tools atau aplikasi rapidminer.

	A	B	C	D	E	F	G	H	I
1	Education	JoiningYear	City	PaymentTier	Age	Gender	EverBenched	ExperienceInCurrentDomain	LeaveOrNot
2	Bachelors	2017	Bangalore	3	34	Male	No	0	0
3	Bachelors	2013	Pune	1	28	Female	No	3	1
4	Bachelors	2014	New Delhi	3	38	Female	No	2	0
5	Masters	2016	Bangalore	3	27	Male	No	5	1
6	Masters	2017	Pune	3	24	Male	Yes	2	1
7	Bachelors	2016	Bangalore	3	22	Male	No	0	0
8	Bachelors	2015	New Delhi	3	38	Male	No	0	0
9	Bachelors	2016	Bangalore	3	34	Female	No	2	1
10	Bachelors	2016	Pune	3	23	Male	No	1	0
11	Masters	2017	New Delhi	2	37	Male	No	2	0
12	Masters	2012	Bangalore	3	27	Male	No	5	1
13	Bachelors	2016	Pune	3	34	Male	No	3	0
14	Bachelors	2018	Pune	3	32	Male	Yes	5	1
15	Bachelors	2016	Bangalore	3	39	Male	No	2	0
16	Bachelors	2012	Bangalore	3	37	Male	No	4	0
17	Bachelors	2017	Bangalore	1	29	Male	No	3	0
18	Bachelors	2014	Bangalore	3	34	Female	No	2	0
19	Bachelors	2014	Pune	3	34	Male	No	4	0
20	Bachelors	2015	Pune	2	30	Female	No	0	1
21	Bachelors	2016	New Delhi	2	22	Female	No	0	1

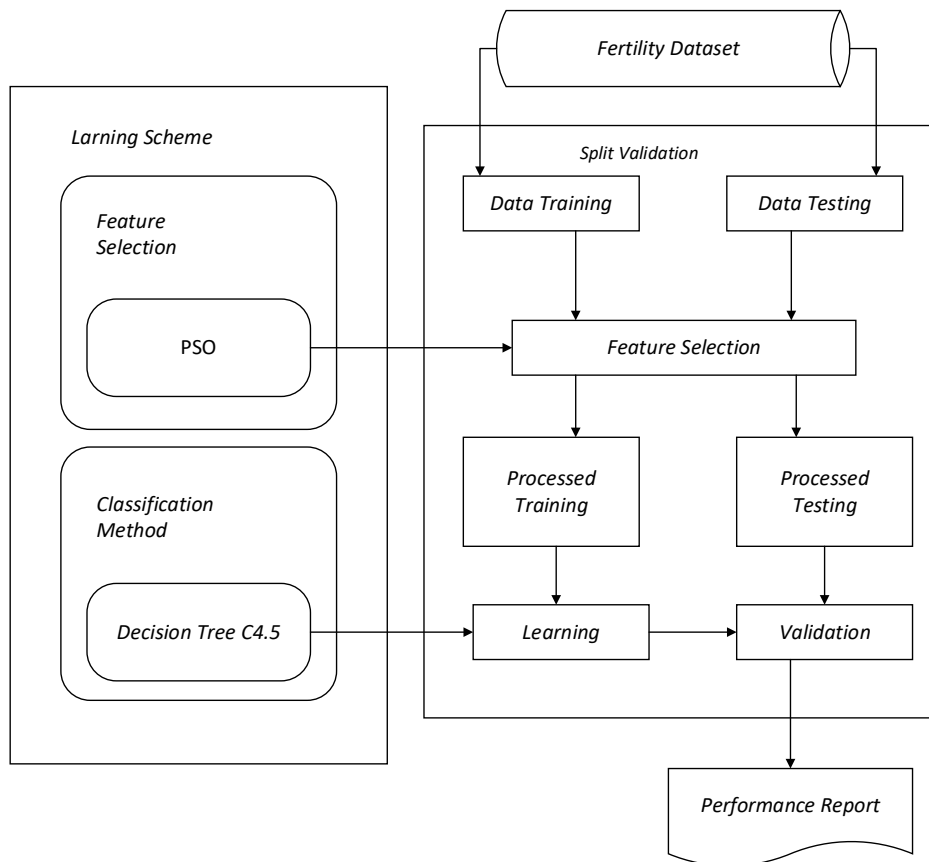
Gambar 3. 4 Dataset Transformation

Semua atribut berjumlah 9, semua akan digunakan dimana terdapat 1 atribut label dan 8 atribut fitur. Penjelasan data dapat dilihat seperti gambar dibawah ini.

Tabel 3. 1 Keterangan Atribut Dataset

Atribut	Deskripsi	Kriteria
Education	Sekolah	Bachelors
		Masters
		PHD
JoiningYear	Tahun Bergabung	2012-2018
City	Kota	Bangalore
		Pune
		New Delhi
PaymentTier	Range Gaji	1= Highest (Tinggi)
		2= Mid Level (Sedang)
		3= Lowest (Rendah)
Age	Umur	22-41
Gender	Jenis Kelamin	Male = Laki-laki
		Female = Perempuan
EverBenched	Tidak pernah keluar dai proyek selama 1 bulan	Yes = Iya
		No = Tidak
ExperienceInCurrentDomain	Pengalaman di lapangan saat ini	0-7 (palingg tinggi 7)
LeaveOrNot	Apakah karyawan tinggalkan perusahaan dalam 2 tahun mendatang	0 = No
		1 = Yes

- Proses mining menggunakan metode *Particle Swam Optimization* (PSO) dan Algoritma *Decision Tree* C4.5 untuk meningkatkan akurasi.



Gambar 3. 5 Alur Proses PSO

- Tahapan pertama yaitu pengumpulan *dataset fertility* yang diambil dari *Kaggle repository*.
- Berikutnya pembagian data menggunakan *split validation*, yaitu membagi data menjadi *data training* dan *data testing*.
- Pada tahap *learning schame* menggunakan *Particle Swarm Optimization* untuk optimasi, sehingga akan mengoptimasi nilai bobot pada setiap atribut.

#### Pengujian PSO

- Selanjutnya proses klasifikasi menggunakan algoritma *Decision Tree C4.5* dan masuk ketahap validasi atau melihat kebenaran dan hasil dari proses sebelumnya.
- Pengujian Algoritma *Decision Tree C4.5*

Pengujian data yang digunakan dalam algoritma C4.5 adalah dataset prediksi loyalitas karyawan yang didapatkan dari Kaggle. Data sampling sebanyak 10 data yang digunakan dalam pengujian ini dapat dilihat pada gambar berikut.

57	PHD	2013	New Delh	3	28	Male	No	2	0
58	Bachelors	2016	Bangalore	3	38	Male	No	2	0
59	Masters	2014	Pune	3	39	Female	No	2	0
60	Bachelors	2013	Pune	3	32	Male	No	4	0
61	Masters	2017	Pune	2	36	Male	No	2	1
62	Masters	2013	New Delh	3	29	Male	No	3	0
63	Bachelors	2015	Bangalore	3	30	Male	No	5	0
64	PHD	2013	Bangalore	2	25	Male	No	3	1
65	Bachelors	2014	Pune	1	22	Female	No	0	1
66	Bachelors	2016	Bangalore	3	34	Male	No	4	0

Gambar 3. 6 Dataset Prediksi Loyalitas Karyawan

Menghitung jumlah karyawan yang keluar dan bertahan, serta entropy dari semua kasus dan semua dibagi berdasarkan atribut education, joining year, city, payment tier, age, gender, ever benched, experience in current domain. Perhitungan nilai entropy total dapat dicari dengan rumus berikut.

$$\begin{aligned}
 \text{Entropy (S)} &= (-3/10) \cdot \log_2(4/10) + (-7/10) \cdot \log_2(7/10) \\
 &= (-0,3) \cdot \log_2(0,3) + (-0,7) \cdot \log_2(0,7) \\
 &= (-0,3) \cdot (-1,736) + (-0,7) \cdot (-0,514) \\
 &= 0,5208 + 0,3598 \\
 &= 0,881291
 \end{aligned}$$

Setelah itu dilakukan perhitungan gain untuk setiap atribut, hasil perhitungan seperti dibawah ini.

Tabel 3. 2 Perhitungan Node Akar

		Jumlah (S)	Yes (Si)	No (Si)	Entropy	Gain
Total		10	3	7	0,881291	
Education						0,044838102
	Bachelors	5	1	4	0,721928	
	Masters	3	1	2	0,918296	
	PHD	2	1	1	1	
JoiningYear						0,356779649
	2012	0	0	0	0	
	2013	4	1	3	0,811278	
	2014	2	1	1	1	
	2015	1	0	0	0	
	2016	2	0	0	0	
	2017	1	1	0	0	
	2018	0	0	0	0	
City						0,156779649
	Bangalore	4	1	3	0,811278	
	Pune	4	2	2	1	
	New Delhi	2	0	0	0	
PaymentTier						0,881290899
	1	1	1	0	0	
	2	2	2	0	0	
	3	7	0	0	0	
Age						0,091277446
	<=29	4	2	2	1	
	>29	6	1	5	0,650022	
Gender						0,0322684
	Male	8	2	6	0,811278	
	Female	2	1	1	1	
EverBenchd						0,00000000
	Yes	0	0	0	0	
	No	10	3	7	0,881291	
ExperienceInCurrentDomain						-0,83007500
	0	1	1	0	0	
	1	0	0	0	0	
	2	4	1	3	0,811278	
	3	2	1	1	1	
	4	2	0	2	0	
	5	1	0	1	0	
	6	0	0	0	0	
	7	0	0	0	0	
Nilai Gain Maksimal						0,88129090

- e. Berikutnya tahapan terakhir yaitu *performance report*, pada tahap akhir ini menampilkan hasil dari pengolahan dari proses sebelumnya. Tahap ini menghasilkan *Confusion Matrix* dan kurva ROC.

5. Tahap interpretasi dan evaluasi adalah mengevaluasi satu atau lebih model yang



disampaikan dalam tahap pemodelan untuk kualitas dan efektivitas sebelum menerapkannya untuk digunakan di lapangan. Pada tahap ini yang akan dilakukan adalah menentukan model sebenarnya mencapai tujuan yang ditetapkan untuk itu pada tahapan pertama, menetapkan apakah beberapa aspek penting dari masalah bisnis atau penelitian belum diperhitungkan secara memadai, mengambil keputusan tentang penggunaan hasil data mining.

### 3.2 Alat dan Bahan

Penelitian ini menggunakan perangkat keras Laptop ASUS VivoBook dengan Processor Intel(R) Core(TM) i5-3470 CPU @ 3.20GHz, RAM 8,00 GB, sedangkan perangkat lunak yang digunakan Microsoft Excel dan RapidMiner Studio Developer 9.6 untuk pengolahan data. Sedangkan bahan yang digunakan diambil dari (<https://www.kaggle.com/datasets/tejashvi14/employee-future-prediction>).

### 3.3 Jadwal Penelitian

Kegiatan	Waktu Kegiatan				
	Oktober	November	Desember	Januari	Februari
Identifikasi Masalah					
Studi Pustaka					
Pengumpulan Data					
Menentukan Metode Data					
Eksperimen dan Hasil					
Penyusunan Laporan					